Kevin Romero Peces-Barba
kromero8@ucsc.edu
1635745

CMPS 142 - Spring 2018

Homework 1 - Problem 2

Let $X = \langle X_1, X_2 \ldots X_n \rangle$ be a vector representing a data instance. Each X_i represents the value of the i^{th} feature for X and can take real values. All X_i's are conditionally independent given the label. We model $P(X1 = x_{1j}|Y = y_k)$ using a Normal distribution as:

$$P(X_1 = x_{1j} | Y = y_k) = \left(\frac{1}{\sigma\sqrt{2\pi}} exp \frac{-(x_{1j} - \mu_{1k})^2}{2\sigma^2} \right)$$

Here $j \in \{1,2,3,...M\}$ represents the j^{th} training instance out of a total of M instances. All instances are iid. Also, x_{1j} refers to the value of the first attribute, X_1 of the j^{th} instance. Note that here we are assuming that all classes, k, have the same variance, σ. In this question, we will estimate the parameter (mean of the Gaussian) for the first attribute. What is the maximum likelihood estimate for $\mu\hat{}_{1k}$? Show your derivation.

We start with the likelihood function for a Normal distribution:

$$L(\mu, \sigma^2 | x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} exp \frac{-(x_1 - \mu)^2}{2\sigma^2} \right) \ldots \left(\frac{1}{\sqrt{2\pi\sigma^2}} exp \frac{-(x_n - \mu)^2}{2\sigma^2} \right) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} exp - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Taking logs...

$$ln\big(L(\mu, \sigma^2 | x)\big) = ln\left(\frac{1}{\sqrt{(2\pi\sigma^2)^n}} exp - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

$$ln\big(L(\mu, \sigma^2 | x)\big) = ln\left(\frac{1}{\sqrt{(2\pi\sigma^2)^n}} \right) + ln\left(exp - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

$$ln\big(L(\mu, \sigma^2 | x)\big) = - ln\left(\sqrt{(2\pi\sigma^2)^n} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$ln\big(L(\mu, \sigma^2 | x)\big) = - \frac{n}{2}\big(ln(2\pi) + ln(\sigma^2)\big) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2¶$$

Now, to calculate MLE, we calculate the derivative with respect to μ:

$$\frac{\delta\mu}{\mu} ln\left(L(\mu, \sigma^2 | x)\right) = 0 - (-2)\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

$$\frac{\delta\mu}{\mu} ln\left(L(\mu, \sigma^2 | x)\right) = \frac{1}{\sigma^2} n(\bar{x} - \mu)$$

If we find the second derivative $\frac{\delta^2\mu}{\mu^2} ln\left(L(\mu, \sigma^2 | x)\right) = -\frac{n}{\sigma^2} < 0$ is negative, so we have a maximum when the first derivative is 0. This happens for $\mu = \bar{x}$.

Going back to μ_{ik}, its MLE is $\bar{x}_1$.